数理解析研究所講究録
916巻 1995年 189-195

The concept of generalized amount of information and non-regular estimation

赤平 昌文 (Masafumi Akahira)

1. はじめに

統計的推測理論においては、Fisher 情報量、Kullback-Leibler 情報量などが良く知られていて、正則な場合すなわち分布に正則条件が仮定された場合に有用であるが、非正則な場合には必ずしも有用とはいえない。そこで非正則な場合にも有効な情報量として類似度 (affinity) を用いて導入されている ([AT91],[L90])。本論ではさらに (Rényi 測度型に) 拡張した一般化情報量について述べ、切断分布族に対して極値統計量と漸近補助統計量から成る統計量の一般化情報量の2次の漸近損失は 0 になることを示す ([A95])。このことはその統計量は2次の漸近十分統計量になることとも符合している ([A91b])。

2. 一般化情報量

標本空間 $(\mathcal{X}, \mathcal{B})$ 上の確率測度 P, Q がある σ-有限測度 μ に関して絶対連続であると仮定する。このとき P, Q の間の情報量を

$$I(P,Q) := -8 \log \int_{\mathcal{X}} \left(\frac{dP}{d\mu} \cdot \frac{dQ}{d\mu} \right)^{1/2} d\mu \tag{2.1}$$

によって定義する ([AT91])。ここで右辺の積分値は類似度 (affinity) と呼ばれている ([M55])。上記の情報量は測度 μ のとり方には依存しない。この情報量は非正則な場合の母数推定問題を考察する際に有用であり ([AT91])、類似度の観点からも検討されている ([L90])。また、この情報量を Rényi 測度型に拡張した一般化情報量を、各 α $(-1 < \alpha < 1)$ について

$$I^{(\alpha)}(P,Q) := -\frac{8}{1-\alpha^2} \log \int_{\mathcal{X}} \left(\frac{dP}{d\mu} \right)^{(1-\alpha)/2} \left(\frac{dQ}{d\mu} \right)^{(1+\alpha)/2} d\mu \tag{2.2}$$

によって定義する ([A95])。この情報量も測度 μ のとり方には依存しない。特に $\alpha = 0$ とすれば一般化情報量 (2.2) は情報量 (2.1) に一致する。すなわち $I^{(0)}(P,Q) = I(P,Q)$ である。

次に $X_1,\ldots,X_n$ をたがいに独立にいずれも (σ-有限測度 μ に関する) 密度関数 $f(x,\theta)$ $(\theta\in\Theta)$ に従う実確率変数とする。ただし Θ は母数空間とし、$\mathbb{R}^1$ の開区間とする。このとき任意の $\theta_1,\theta_2\in\Theta$ に対して $f(\cdot,\theta_1)$ と $f(\cdot,\theta_2)$ の間の X_1 に関する一般化情報量 (generalized amount of information) を $-1<\alpha<1$ なる各 α について

$$I_{X_1}^{(\alpha)}(\theta_1,\theta_2) := -\frac{8}{1-\alpha^2}\log\int_{-\infty}^{\infty} f(x,\theta_1)^{(1-\alpha)/2} f(x,\theta_2)^{(1+\alpha)/2} d\mu(x)$$

で表わす。同様にして $f(\cdot,\theta_1)$ と $f(\cdot,\theta_2)$ の間の $\mathbf{X}=(X_1,\ldots,X_n)$ に関する一般化情報量を $I_{\mathbf{X}}^{(\alpha)}(\theta_1,\theta_2)$ で表わせば

$$I_{\mathbf{X}}^{(\alpha)}(\theta_1,\theta_2) = nI_{X_1}^{(\alpha)}(\theta_1,\theta_2)\ \ (-1<\alpha<1) \tag{2.3}$$

になる。もっと一般に、統計量 $T_n := T_n(\mathbf{X})$ の密度関数 $g(t,\theta)$ が与えられれば、同様にして T_n に関する一般化情報量も定義できて、それを $I_{T_n}^{(\alpha)}(\cdot,\cdot)$ で表わす。このとき適当な正則条件の下では

$$I_{T_n}^{(\alpha)}(\theta_1,\theta_2) \le I_{\mathbf{X}}^{(\alpha)}(\theta_1,\theta_2) \qquad (-1<\alpha<1)$$

が成り立つ。そこで $-1<\alpha<1$ なる各 α について統計量 T_n の一般化情報量損失

$$I_{\mathbf{X}}^{(\alpha)}(\theta_1,\theta_2) - I_{T_n}^{(\alpha)}(\theta_1,\theta_2)$$

を考える。本論では $|\theta_1-\theta_2| = O(n^{-1})$ のときにその一般化情報量損失を $o(n^{-1})$ の次数まで考える。

ここで一般化情報量と Fisher 情報量の関係について考える。適当な正則条件の下で、任意の α $(-1<\alpha<1)$ と十分小さい $\Delta\theta$ について次のことが成り立つ。

$$\begin{aligned}
I_{X_1}^{(\alpha)}(\theta,\theta+\Delta\theta) &= -\frac{8}{1-\alpha^2}\log\int_{-\infty}^{\infty} f(x,\theta)^{(1-\alpha)/2} f(x,\theta+\Delta\theta)^{(1+\alpha)/2} d\mu(x) \\
&= -\frac{8}{1-\alpha^2}\log\int_{-\infty}^{\infty} \exp\{\frac{1-\alpha}{2}\log f(x,\theta) + \frac{1+\alpha}{2}\log f(x,\theta+\Delta\theta)\} d\mu(x) \\
&= -\frac{8}{1-\alpha^2}\log\Big[1 + \frac{(1+\alpha)(\Delta\theta)^2}{4}\int_{-\infty}^{\infty}\frac{\partial^2\log f(x,\theta)}{\partial\theta^2} f(x,\theta)d\mu(x) \\
&\qquad + \frac{(1+\alpha)^2(\Delta\theta)^2}{8}\int_{-\infty}^{\infty}\{\frac{\partial\log f(x,\theta)}{\partial\theta}\}^2 f(x,\theta)d\mu(x) + o((\Delta\theta)^2)\Big] \\
&= -\frac{8}{1-\alpha^2}\log\Big[1 - \frac{(1-\alpha^2)(\Delta\theta)^2}{8} I_{X_1}(\theta) + o((\Delta\theta)^2)\Big] \\
&= I_{X_1}(\theta)(\Delta\theta)^2 + o((\Delta\theta)^2)
\end{aligned}$$

ただし $I_{X_1}(\theta) = E_\theta[\{(\partial/\partial\theta)\log f(X_1,\theta)\}^2]$ (Fisher 情報量) とする。

3. 一般化情報量損失

$X_1, \ldots, X_n$ をたがいに独立にいずれも (ルベーグ測度に関して) 密度関数 $f(x, \theta)$ $(\theta \in \Theta)$ に従う実確率変数とする。このとき θ が位置母数、すなわち $f(x, \theta) = f_0(x - \theta)$ の場合を考える。さらに次の条件を仮定する。

(A.1) $f_0(x) > 0 \ \ (a < x < b)$; $f_0(x) = 0 \ \ (x \leq a, x \geq b)$.
ただし a, b は有限とする。

(A.2) $f_0(x)$ は開区間 (a, b) において 2 回連続微分可能で、

$$\lim_{x \to a+0} f_0(x) = \lim_{x \to b-0} f_0(x) = c,$$

$$\lim\nolimits_{x \to b-0} f_0'(x) = -\lim\nolimits_{x \to a+0} f_0'(x) = h$$

である。ただし c は正の定数で、h は定数とする。

(A.3) $0 < I_0 := \int_a^b \{f_0'(x)\}^2 / f_0(x) dx < \infty.$

ここで

$$I = -\int_a^b \frac{d^2 \log f_0(x)}{dx^2} f_0(x) dx$$

とおくと、条件 (A.1) ∼ (A.3) より

$$I - I_0 = -2h \tag{3.1}$$

になる。上記のような設定の下では、一致性の order は n であることが知られている。このとき X_1 および $\mathbf{X}$ に関する一般化情報量は次のようになる。

定理 3.1. 条件 (A.1) ∼ (A.3) を仮定する。このとき任意の α $(-1 < \alpha < 1)$ と十分小さい Δ に対して

$$\begin{aligned} I_{X_1}^{(\alpha)}(\theta, \theta + \Delta) &= \frac{1}{1 - \alpha^2}[8c|\Delta| + \{4c^2 - 2h + I - \alpha^2(2h + I)\}\Delta^2] + o(\Delta^2), \\ I_{\mathbf{X}}^{(\alpha)}(\theta, \theta + \Delta) &= \frac{1}{1 - \alpha^2}[8cn|\Delta| + \{4c^2 - 2h + I - \alpha^2(2h + I)\}n\Delta^2] + o(n\Delta^2) \end{aligned}$$

が成り立つ。

証明の概略. $\theta = 0$ として一般性を失わない。$\Delta > 0$ とする。また $l(x) = \log f_0(x)$ とおき、$-1 < \alpha < 1$ なる α を任意に固定する。このとき

$$f_0(x)^{(1-\alpha)/2} f_0(x - \Delta)^{(1+\alpha)/2}$$

$$= f_0(x)\left[1 - \frac{1+\alpha}{2}\Delta l'(x) + \frac{1+\alpha}{4}\Delta^2 l''(x) + \frac{(1+\alpha)^2}{8}\Delta^2\{l'(x)\}^2 + o(\Delta^2)\right]$$

より

$$\int_{a+\Delta}^{b} f_0(x)^{(1-\alpha)/2} f_0(x-\Delta)^{(1+\alpha)/2} dx$$

$$= \int_{a+\Delta}^{b} f_0(x)dx - \frac{1+\alpha}{2}\Delta\int_{a+\Delta}^{b} l'(x)f_0(x)dx + \frac{1+\alpha}{4}\Delta^2\int_{a+\Delta}^{b} l''(x)f_0(x)dx$$

$$+\frac{(1+\alpha)^2}{8}\Delta^2\int_{a+\Delta}^{b}\{l'(x)\}^2 f_0(x)dx + o(\Delta^2)$$

$$= 1 - c\Delta - \frac{\alpha}{2}h\Delta^2 - \frac{1+\alpha}{4}I\Delta^2 + \frac{(1+\alpha)^2}{8}(I+2h)\Delta^2 + o(\Delta^2)$$

になる。よって $-1<\alpha<1$ に対して

$$I_{X_1}^{(\alpha)}(0,\Delta) = \frac{8}{1-\alpha^2}c\Delta + \frac{1}{1-\alpha^2}\{4c^2 - 2h + I - \alpha^2(2h+I)\}\Delta^2 + o(\Delta^2)$$

となる。$\Delta<0$ のときも同様に得られる。また (2.3) から $I_{\mathbf{X}}^{(\alpha)}(0,\Delta)$ も得られる。 ■

次に極値統計量 $\overline{\theta}$ と $\underline{\theta}$ を

$$\overline{\theta} := \min_{1\le i\le n} X_i - a, \quad \underline{\theta} := \max_{1\le i\le n} X_i - b$$

とし、

$$Z_1(\theta) := -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{f_0'(X_i-\theta)}{f_0(X_i-\theta)} \quad (\underline{\theta}<\theta<\overline{\theta})$$

とする。また、$\hat{\theta}^* = (\underline{\theta}+\overline{\theta})/2$ とおくと、$\hat{\theta}^*$ は θ の一致推定量になる。このとき $Z_1^* := Z_1(\hat{\theta}^*)$ とすると、Z_1^* は漸近補助統計量になる。さらに $U = n(\overline{\theta}-\theta)$, $V = n(\underline{\theta}-\theta)$ とおくと (U,V) の漸近同時密度 $g_n(u,v)$ は

$$g_n(u,v) = \begin{cases} c^2 e^{-c(u-v)}[1 + \frac{1}{n}\{-1 + 2c(u-v) + \frac{h}{4}((u+v)^2 + (u-v)^2) \\ \qquad\qquad -\frac{c^2}{2}(u-v)^2 - \frac{h}{c}(u-v)\}] + o\left(\frac{1}{n}\right) & (v<0<u), \\ 0 & (\text{その他}) \end{cases} \tag{3.2}$$

になる ([A91a])。そこで統計量 $T_n^* := (Z_1^*/\sqrt{n}I_0, \overline{\theta}, \underline{\theta})$ の一般化情報量を求める。

定理 3.2. 条件 (A.1) ~ (A.3) を仮定する。このとき $\Delta = O(1/n)$ とすれば、統計量 T_n^* の一般化情報量は、任意の α $(-1 < \alpha < 1)$ に対して

$$I_{T_n^*}^{(\alpha)}(\theta, \theta+\Delta) = \frac{8}{1-\alpha^2} cn|\Delta| + \left\{\frac{4}{1-\alpha^2}(c^2-h) + I_0\right\} n\Delta^2 + o(n\Delta^2)$$

である。

証明の概略. $\theta = 0$ として一般性を失わない。α を区間 $(-1, 1)$ に固定する。まず

$$I_{T_n^*}^{(\alpha)}(0, \Delta) = -\frac{8}{1-\alpha^2} \log E[\exp\{-\frac{1-\alpha^2}{8} I_{Z_1^*/(\sqrt{n}I_0)|\underline{\theta},\overline{\theta}}^{(\alpha)}(0, \Delta)] + I_{\underline{\theta},\overline{\theta}}^{(\alpha)}(0, \Delta) \tag{3.3}$$

であることに注意する。ここで $I_{Z_1^*/(\sqrt{n}I_0)|\underline{\theta},\overline{\theta}}^{(\alpha)}(0, \Delta)$ は $\underline{\theta}$, $\overline{\theta}$ が与えられたときの $Z_1^*/(\sqrt{n}I_0)$ の条件付分布に関する一般化情報量とする。次に (3.2) より

$$\begin{aligned}&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_n(u, v)^{(1-\alpha)/2} g_n(u-n\Delta, v-n\Delta)^{(1+\alpha)/2} dudv \\ &= e^{-2cn|\Delta|}\{1 + (h-c^2)n\Delta^2 + o(n\Delta^2)\}\end{aligned}$$

となるから

$$I_{\underline{\theta},\overline{\theta}}^{(\alpha)}(0, \Delta) = \frac{8}{1-\alpha^2} cn|\Delta| + \frac{4}{1-\alpha^2}(c^2-h)n\Delta^2 + o(n\Delta^2) \tag{3.4}$$

を得る。また

$$I_{Z_1^*/(\sqrt{n}I_0)|\underline{\theta},\overline{\theta}}^{(\alpha)}(0, \Delta) = I_0 n\Delta^2 + o(n\Delta^2)$$

であるから、(3.3), (3.4) より

$$I_{T_n^*}^{(\alpha)}(0, \Delta) = \frac{8}{1-\alpha^2} cn|\Delta| + \left\{\frac{4}{1-\alpha^2}(c^2-h) + I_0\right\} n\Delta^2 + o(n\Delta^2)$$

を得る。 ■

次に、統計量 $T_n := T_n(\mathbf{X})$ の 2 次の一般化情報量損失を、$-1 < \alpha < 1$ となる α に対して

$$L_n^{(\alpha)}(T_n) := \frac{1}{n\Delta^2}\left\{I_{\mathbf{X}}^{(\alpha)}(\theta, \theta+\Delta) - I_{T_n}^{(\alpha)}(\theta, \theta+\Delta)\right\} + o(1)$$

で定義する。ただし $\Delta = O(1/n)$ とする。このとき次のことが成り立つ。

定理 3.3. 条件 (A.1) ~ (A.3) を仮定する。このとき $\Delta = O(1/n)$ とすれば、統計量 T_n^* の 2 次の一般化情報量損失は、任意の α $(-1 < \alpha < 1)$ に対して

$$L_n^{(\alpha)}(T_n^*) = o(1)$$

である。

証明は (3.1) と定理 3.1, 3.2 から得られる。定理 3.3 から統計量 T_n^* の 2 次の一般化情報量損失は 0 になり、これは α には無関係である。従って上記の結果はこのような型の一般化情報量に関して不変になる。また [A91b] において、一方向型分布族に対して T_n^* が 2 次の漸近十分統計量になることが示されているが、そのことは定理 3.3 の結果とも合致している。さらに、T_n^* に基づく推定量として Weiss and Wolfowitz [WW67] の最大確率推定量 (maximum probability estimator) があり、これは $h=0$ のとき局所的に漸近有効推定量になる ([A91a])。

例 (切断正規分布の場合). 確率変数 $X_1,\dots,X_n$ をたがいに独立にいずれも密度関数

$$f_0(x-\theta)=\begin{cases} ce^{-(x-\theta)^2/2} & (|x-\theta|<1) \\ 0 & (\text{その他}) \end{cases}$$

をもつ分布に従うとする。ただし θ は実母数で、c はある正の定数とする。このとき

$$\lim_{x\to-1+0} f_0(x)=\lim_{x\to1-0} f_0(x)=ce^{-1/2},$$

$$h=\lim_{x\to1-0} f_0'(x)=-\lim_{x\to-1+0} f_0'(x)=-ce^{-1/2}$$

になる。従って条件 (A.1) $\sim$ (A.3) は満たされるから、定理 3.3 より統計量 $(Z_1^*/(\sqrt{n}I_0),\overline{\theta},\underline{\theta})$ の 2 次の一般化情報量損失は 0 になる。ここで $Z_1^*=\sqrt{n}(\overline{X}-\hat{\theta}^*)$, $\overline{X}=X_i/n$, $\hat{\theta}^*=(\overline{\theta}+\underline{\theta})/2$, $\underline{\theta}=\max_{1\le i\le n}X_i-1$, $\overline{\theta}=\min_{1\le i\le n}X_i+1$, $I_0=1-2ce^{-1/2}$ である。

参考文献

[A91a] Akahira, M. (1991). The 3/2th and 2nd order asymptotic efficiency of maximum probability estimators in non-regular cases. *Ann. Inst. Statist. Math.*, **43**, 181–195.

[A91b] Akahira, M. (1991). Second order asymptotic sufficiency for a family of distributions with one-directionality. *Metron* **49**, 133–143.

[A95] Akahira, M. (1995). Loss of information of a statistic for a family of non-regular distributions. To appear in *Ann. Inst. Math. Statist.*

[AT91] Akahira, M. and Takeuchi, K. (1991). A definition of information amount applicable to non-regular cases. *Journal of Computing and Information* **2**, 71–92.

[L90] Le Cam, L. (1990). On standard asymptotic confidence ellipsoids of Wald. *Internat. Statist. Rev.*, **58**, 129–152.

[M55] Matusita, K. (1955). Decision rules based on the distance for problems of fit, two samples and estimation. *Ann. Math. Statist.*, **26**, 631–640.

[WW67] Weiss, L. and Wolfowitz, J. (1967). Maximum probability estimators. *Ann. Inst. Statist. Math.*, **19**, 193–206.